ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 40,No. 3,May 2025,pp. 686－698
DOI:10. 16337/j. 1004-9037. 2025. 03. 010

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于双重对比学习模型的SAR自动目标识别背景去偏方法

张文青[1]，王　景[1]，黄雪琴[1]，田巳睿[1]，何　成[2]，张劲东[3]，李洪涛[1]

(1. 南京理工大学电子工程与光电技术学院,南京 210094;2. 北京遥感设备研究所,北京 100006;3. 南京航空航天大学电子信息工程学院 ,南京 211106)

摘　要：对比学习作为一种自监督方法,可从无标记SAR图像中挖掘目标表征,是SAR自动目标识别(Automatic target recognition, ATR)的关键技术。但现有模型常将目标与背景整体表征,导致特征混杂背景干扰,从而削弱模型对目标的聚焦能力。为解决这一问题,提出了一种多分支双重对比学习模型。该模型在保留传统实例对比分支的基础上,创新性引入背景纠偏对比分支,构建了多分支对比学习框架;通过正负样本中目标与背景的随机组合策略,并结合ResNet50的主干网络以及自注意力池化增强语义特征提取,利用优化的双重对比损失函数改进目标特征学习,降低背景与目标的伪相关性;基于MSTAR数据集的Shapley值分析验证了该模型的有效性,目标分类结果证明该方法显著增强了模型特征提取的因果性,大大提升了SAR ATR算法的泛化性能。

关键词：SAR自动目标识别;自监督对比学习;表征学习;背景去偏

中图分类号：TN958;TN957.52　　**文献标志码**：A

Dual Contrastive Learning Model Based Background Debiasing in SAR ATR

ZHANG Wenqing[1], WANG Jing[1], HUANG Xueqin[1], TIAN Sirui[1], HE Cheng[2], ZHANG Jingdong[3], LI Hongtao[1]

(1. School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; 2. Beijing Institute of Remote Sensing Equipment, Beijing 100006, China; 3. College of Electronic and Information Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China)

Abstract: Contrastive learning, as a self-supervised approach, enables the extraction of target representations from unlabeled SAR images, serving as a critical technique for automatic target recognition (ATR) in SAR. However, existing models often encode targets and backgrounds holistically, resulting in feature representations contaminated by background interference, which diminishes the model's ability to focus on targets. To address this issue, this paper proposes a novel multi-branch dual contrastive learning model. Firstly, the model retains the conventional instance contrastive branch while introducing an innovative background correction contrastive branch, establishing a multi-branch contrastive learning framework. Secondly, through a random recombination strategy of targets and backgrounds in positive and negative samples, combined with the ResNet50 backbone network and self-attention pooling to enhance semantic feature extraction, an optimized dual contrastive loss function is employed to refine target feature learning and mitigate spurious correlations between backgrounds and targets. Finally, Shapley value

基金项目:国家自然科学基金重点项目(41930110);国家自然科学基金(62171224)。
收稿日期:2024-06-02;**修订日期**:2024-10-22

analysis based on the MSTAR dataset validates the model's effectiveness, and target classification results demonstrate that this approach significantly enhances the causality of feature extraction, substantially improving the generalization performance of SAR ATR algorithms.

Key words: SAR ATR; self-supervised contrastive learning; representation learning; background debiasing

引　　言

在SAR自动目标识别(Automatic target recognition,ATR)领域中,表征学习是决定目标识别性能的关键技术之一[1]。而对比学习作为一种高效的自监督表征学习方法,因其出色的特征提取能力被广泛应用[2]。然而,基于对比学习的模型通常将包含目标和背景的整个切片输入深度网络进行学习,导致提取的特征不仅反映目标特性,还包含对分类任务无效的背景信息[3]。同时由于SAR ATR训练数据集数据往往在相似背景条件下获取的采集特性,使得模型易对背景信息产生过度依赖。尽管背景与目标类别无直接因果关系,但背景变化仍会削弱模型的泛化能力,影响下游分类性能。其中,以MSTAR数据集为例[4],其目标类型与背景杂波高度相关[5],研究表明,即便去除目标区域,仅利用背景杂波特征,传统分类器的识别准确率仍可超过99%[6]。然而,在实际应用中,目标所处的环境可能千差万别,因此SAR图像中的背景杂波特性及其分布也将完全不同,从而会影响实际装备的分类性能和作战能力。因此,表征学习模型需具备从目标切片中提取目标固有特性并抑制背景偏差的能力,以增强SAR ATR系统的泛化能力,应对环境变化挑战。

为实现对背景偏差的抑制,首先需分析目标切片中不同区域对下游识别任务的贡献,并根据贡献差异优化表征学习模型。现有可解释性分析方法多聚焦于像素级贡献,难以定量评估区域级影响。为此,研究人员采用基于Shapley值[7]的显著性图分析法,量化各区域对识别性能的贡献,为模型因果关系提供明确指标[8]。Belloni等[9]基于此方法发现,在SAR图像中加入背景杂波可显著提高模型的分类性能,李玮杰等[10]基于Shapley值分析了揭示训练中目标与背景区域的贡献及相互作用。因此,本文也基于Shapley值的显著性图方法,将具有高效端到端结构与卓越分类性能[11]的SimCLR模型作为表征学习的基础模型,量化其在训练期间目标区域和背景区域对识别结果的贡献,结果如图1所示。随着训练轮次的增加,目标区域和背景区域的Shapley值均呈现上升趋势。这表明模型在识别目标时充分利用了所有区域的信息,同时也反映出所学特征不仅与目标区域相关,还与背景区域存在伪相关,进而影响模型的泛化能力。这些研究进一步表明,深度神经网络所学到的特征不仅包含了分类所需的目标信

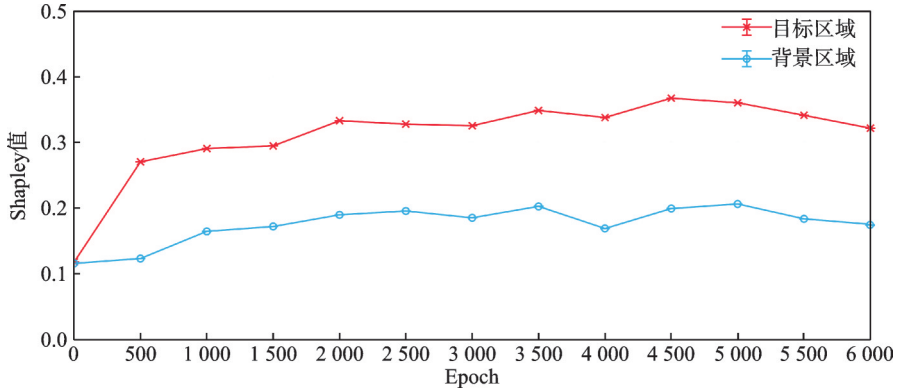


图1　SimCLR模型训练过程中SAR图像目标区域及背景区域的Shapley值曲线

Fig.1　Shapley value curves of SAR target area and background area during SimCLR model training process

息,还混杂了背景的非因果伪相关信息,导致特征对背景变化的泛化能力不足,进而降低下游任务性能。

针对这一伪相关性问题,近年研究提出了多种抑制背景影响的策略。Heiligers等[12]将图像分割与深度学习结合,仅输入目标区域至网络,以降低背景对分类的影响,但略牺牲准确率。Li等[13]融合自动编码器与目标提取网络,自动学习目标特征,减少背景噪声干扰。Wang等[14]在传统卷积神经网络中引入压缩与激发(Squeeze and excitation,SE)模块,通过特征权重分配增强对非因果背景的抑制。Feng等[15]提出一种基于属性散射中心(Attributed scattering center,ASC)模型的目标部分注意网络(Part attention network,PAN),融合目标部分特征,削弱背景表达。Lewis等[16]采用数据增强和域对齐提取鲁棒特征。侯兴荣等[17]通过挖掘全局与局部结构信息,提升类内相似性与类间差异,增强鲁棒性。Chen等[11]提出通过信号杂波比(Signal-to-clutter ratio,SCR)对样本进行重新加权,以减小不同类别样本间的SCR差异,抑制背景伪相关。Peng等[18]提出了对比特征对齐(Contrastive feature alignment,CFA)方法,通过对比调整高级特征与背景关系,学习可靠目标表征。此外,Li等[19]施加稀疏约束并结合对比损失,利用辅助掩码分离目标特征并进行域对齐。然而,这些方法对表征学习模型的性能优化不足,部分仅适用于特定数据集,缺乏普适性。因此,开发一种能够精准提取目标特征、隔离背景信息并消除伪相关性对识别性能负面影响的方法,仍是当前自监督表征学习的核心挑战。

为应对上述挑战,本文创新性地提出了一种基于多分支结构的背景去偏双重对比学习(Dual contrastive learning for background debiasing,BDDCL)模型,旨在解决SAR自监督表征学习中由数据选择偏差引发的背景区域与目标类别之间的伪相关问题。首先,选用SimCLR作为基础模型,独创性地引入背景纠偏对比分支,构建了多分支对比学习框架,以解耦背景与目标信息。其次,通过精心设计的正样本不共享背景、负样本共享相似背景的目标与背景随机组合策略,并结合残差网络(Residual network 50,ResNet50)的主干网络以及自注意力池化增强语义特征提取,并利用优化的双重对比损失函数改进目标特征学习,有效降低伪相关性并提升整体识别性能。最后,通过MSTAR数据集实验与Shapley值分析,验证模型在特征因果性与分类精度方面的改进,充分展示其在复杂场景下的泛化能力与鲁棒性提升。

1 基于双重对比学习模型的背景伪相关去偏方法

1.1 双重对比学习模型框架

为了解决SAR ATR任务中目标类别与背景之间可能产生的伪相关偏差问题,本文基于SimCLR模型,提出了一种基于多分支结构的背景去偏双重对比学习模型。该模型通过在特征空间中拉近同类目标特征之间的距离,同时区分背景相似但目标不同的样本,从而有效消除背景伪相关的影响,提升模型的泛化能力。

如图2所示,BDDCL模型采用多分支结构,即传统分支L_{con}^{1}以及新增分支L_{con}^{2}和L_{con}^{3}。其中,在传统分支L_{con}^{1}中,沿用SimCLR对比学习优势,利用信息噪声对比估计(Information noise-contrastive estimation,InfoNCE)实例级判别损失[20]学习样本特征。具体而言,首先将两种图像增强方式t_1: T和t_3: T'作用于样本x上,生成同一样本的不同增强视图V和V',对于锚点样本V,V'为其正样本,而来自其他图像的增强视图则被视为与之不相似的负样本。然后,模型通过实例级损失函数优化,即通过最小化锚点样本与其正样本之间的距离,同时最大化与负样本的距离来进行学习。这种方法促使同一实例的不同视图特征彼此靠近,而不同实例的特征则相互排斥。然而,在传统分支L_{con}^{1}中,正样本通常来自于同一实例的不同增强图像(如V和V'),负样本是来自不同实例的增强数据。这种样本构造策略可能导致正样本之间具有相似的背景信息,进而使模型过度关注于背景特征;与此同时,负样本间背景的显著

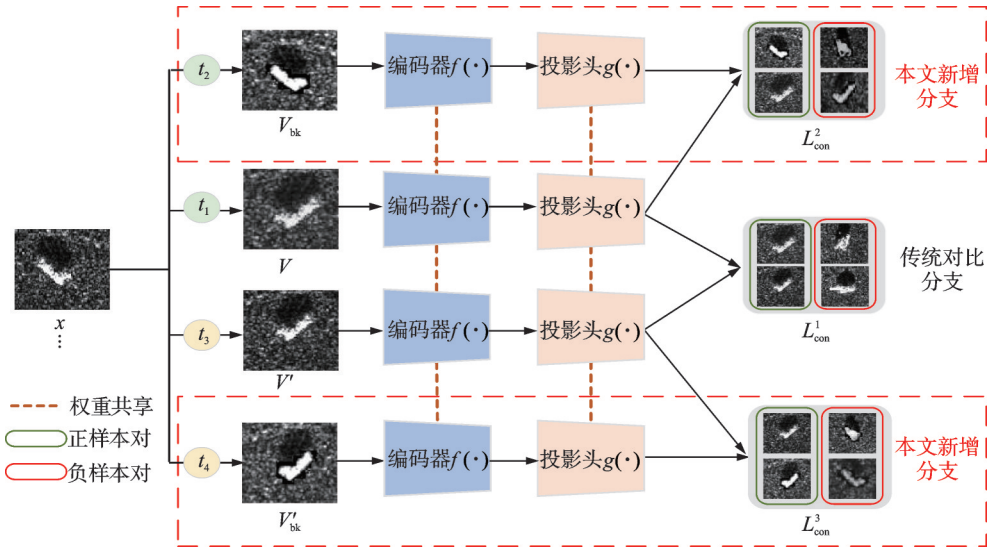


图2 BDDCL模型框架

Fig.2 BDDCL model framework

差异可能使得模型在最小化负样本间相似性时对背景特征过度敏感,降低了模型对背景变化的泛化能力。

为解决这些问题,BDDCL模型创新性新增L^2_{con}和L^3_{con}分支,通过创建与锚点样本背景区域显著不同的样本,消除背景的伪相关性。具体而言,在新增分支L^2_{con}和L^3_{con}中,替换锚点样本背景为随机异类背景,使正样本不再共享背景信息;负样本则与锚点共享背景但目标类别不同,通过替换正样本对中的背景信息,减少背景相关性对模型的干扰。该方法分两步实施:第一步是目标区域与背景区域的分割。通过瑞利双参数恒虚警(Constant false alarm rate,CFAR)检测算法[21],将SAR图像的目标区域与背景区域分离;第二步是构造目标区域增强视图与背景相关抑制视图。首先,目标区域增强视图主要通过单独对锚点样本的目标区域切片施加旋转、剪切、缩放等常规的图像增强t_2:T和t_4:T',从而生成目标区域的增强视图。然后,这些增强视图被置入不同类别的背景区域中,生成对应的正样本对V_{bk}和V'_{bk}。通过这种方式,构造出的正样本在目标信息上相似,均来自同一类别,但背景信息存在显著差异。其次,背景相关抑制是将其他样本的增强目标区域放置在锚点样本的原始背景区域中,从而生成负样本对。这些负样本具有不同的目标信息,但共享相同的背景信息。这种方法通过构建具有不同背景的正样本对和共享相同背景的负样本对,有效减少了目标与背景一致性带来的相关性影响。同时,为了确保正样本之间的相似性仅来源于图像本身,每一对比分支都采用相同的数据增强策略。此外,所有的增强视图在进行对比损失评估以表征样本相似性之前,均通过权重共享的编码器和投影头进行编码,这有助于确保特征表达的一致性并提升模型的表征学习能力。

1.2 基于ResNet结构的主干网络

如图3所示,在实验中,首先基于1.1节提出的正负样本构造方式,生成同一实例的多个增强视图v_i。其次,采用ResNet50架构[22]作为编码器主干网络$f_\theta(\cdot)$,利用其残差连接机制提升特征提取能力,将增强数据映射到特征嵌入空间中,以提取表征投影向量$\boldsymbol{h}_i$。为了降低计算复杂度,模型在ResNet50模块前添加额外的3×3卷积层,用于调整输入图像尺寸。

该模型首先通过输入层接收数据,随后经过多层卷积和池化操作进行特征提取与降维。网络中包

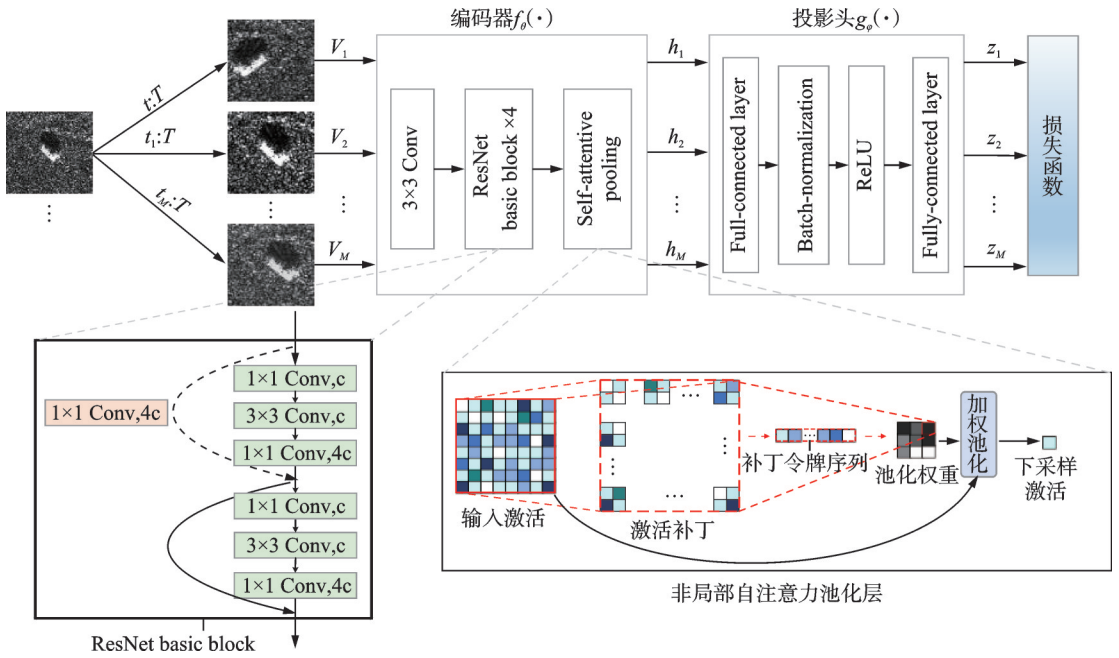


图3 基于ResNet结构的主干网络框架

Fig.3 Backbone network framework based on ResNet structure

含了多个残差模块，旨在通过数个相连的卷积层和残差连接来进一步增强特征学习能力，改善训练效果。在网络的末端，模型引入了自注意力池化层以处理最后一个残差块的输出，将其映射为一个固定长度的一维特征向量$\boldsymbol{h}_i$，完成特征的提取。接着，通过由两个全连接层、归一化层和激活层组成的投影头$g_\varphi(\cdot)$将高维表征向量$\boldsymbol{h}_i$映射到一个具有更好分类性能的低维特征空间，得到用于损失函数计算的编码向量$\boldsymbol{z}_i$。模型以改进的双对比信息最大化(Information maximization，InfoMax)损失为目标函数，通过反向传播算法优化模型并更新模型参数。

为进一步降低计算开销，正样本通过对锚点样本背景进行变换生成增强视图。在目标与背景解耦后，正样本根据背景类别标签编码，并存入负样本字典的对应特征队列中，这些队列随后被用于计算其他锚点样本的损失函数。在训练过程中，负样本从已生成的正样本中基于背景标签采样。负样本字典的使用不仅减少了额外计算开销，还确保了负样本背景信息的一致性。训练完成后，所得到的编码器$f_\theta(\cdot)$和表征嵌入$\boldsymbol{h}_i$将被用于下游的分类任务。

1.3 基于自注意力池化的表征提取与降维

ResNet50网络在卷积层末端采用全局平均池化层生成特征向量。然而，该方法聚焦于局部特征聚合，可能导致信息丢失，尤其对于结构复杂的对象，其难以捕捉整体结构及部分间的关联。为此，将全局平均池化层替换为自注意力池化层，以增强网络对全局关联性的建模能力，提升特征表达效果。

图4显示了该池化模块的整体结构。该模块利用多头自注意模块[23]捕捉不同位置特征点之间的依赖关系，并通过全局信息对编码进行加权。首先，通过补丁嵌入模块来压缩和编码输入$\boldsymbol{x}\in\mathbf{R}^{h\times w\times c_x}$的局部信息，输出令牌序列$\boldsymbol{x}_p\in\mathbf{R}^{\left(\frac{h\times w}{\epsilon_p^2}\right)\times(\epsilon_r c_x)}$，其中$\epsilon_p$为补丁的大小，$\epsilon_r c_x$表示输出通道数。接着，加入可学习的位置编码到$\boldsymbol{x}_p$中以丰富序列信息。随后，采用多头自注意力模块模拟补丁令牌间的长期依赖关

系,生成自注意力的令牌序列x_{attn},且其尺寸不变,。即$\left(b\times\frac{h\times w}{\varepsilon_p^2}\times\varepsilon_r c_x\right)$,其中$b$表示批量大小。该模块中的3个权重矩阵分别为$\boldsymbol{Q}$、$\boldsymbol{K}$和$\boldsymbol{V}$。之后,空间-通道恢复模块解码来自$x_{\text{attn}}$的空间和通道信息,并将其恢复为与输入$\boldsymbol{x}$相同的大小。具体而言,令牌序列首先被重新变形为$\left(b\times\varepsilon_r c_x\times\frac{h}{\varepsilon_p}\times\frac{w}{\varepsilon_p}\right)$,然后通过双线性插值上采样将其扩展到原始空间分辨率($b\times\varepsilon_r c_x\times h\times w$),再通过卷积调整通道数至($b\times c_x\times h\times w$)。最后,在加权池化模块中得到从$x_r$下采样的输出特征映射$\pi(x)$,该特征映射作为权重被赋予到输入激活$x$中,从而生成最终的目标低维表征。这种精心设计的全局视图增强了模型捕捉样本长程依赖关系和类内一致性特征的能力,使其能够更加有效地提取数据特征。

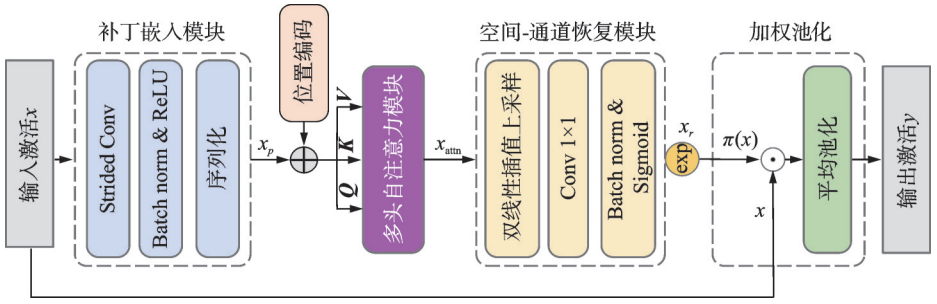


图4 非局部自注意力池化层

Fig.4 Non-local self-attention pooling layer

1.4 对比损失函数设计

利用多分支对比学习构建一个能抑制背景表征偏差的特征空间,并基于多分支的结构设计了本文的对比损失函数

$$\mathcal{L}=(\mathcal{L}_{\text{con}}^1+\mathcal{L}_{\text{con}}^2+\mathcal{L}_{\text{con}}^3)/3 \tag{1}$$

式中:L_{con}^1、L_{con}^2和L_{con}^3分别代表各自分支的损失函数并基于对比学习模型中常用的InfoNCE对比损失[20]

$$L_{\text{con}}=-\log\frac{\exp(\text{sim}_{i,j}/\tau)}{\exp(\text{sim}_{i,j}/\tau)+\sum_{k=1}^{A}\exp(\text{sim}_{i,k}/\tau)} \tag{2}$$

式中:$\text{sim}_{i,j}=\frac{\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{z}_j}{\left(\|\boldsymbol{z}_i\|\cdot\|\boldsymbol{z}_j\|\right)}$,$z=g_\varphi(f_\theta(V))$,$\boldsymbol{z}_i$和$\boldsymbol{z}_j$代表一对正样本的增强视图的表征投影,$A$为负样本的数量,$f_\theta$表示编码器网络,$g_\varphi$代表投影头。对比学习模型的强度由温度系数$\tau$调制。

通过优化上述损失函数,模型有效增强了正样本的相似性,并降低了负样本的相似性,从而强化了对目标区域语义信息的关注,抑制了从非因果背景中提取特征的可能性。因此,同类目标前景的属性在特征空间中会彼此靠近,而不相关的背景偏差信息则会与锚点样本相互分离。经过训练后,模型能够准确区分目标,仅从目标区域中学习与类别密切相关的稳健表征。BDDCL模型显著减轻了SAR ATR任务中背景偏差的影响,构建出一个对背景变化高度鲁棒的表征空间。

1.5 基于Shapley值的表征贡献度

为评估本文模型性能改进,通过Shapley值[7]定量分析输入图像中目标与背景区域对表征性能及识别率的影响。Shapley值源于合作博弈论,可有效衡量深度学习模型中各输入特征的贡献。在MSTAR

数据集上，利用Shapley值量化目标区域和背景区域对分类任务的贡献，阐明其与目标类别的因果关系。

Shapley值是用于计算合作博弈中各参与者对联盟收益的贡献，其主要目标是基于每个参与者对于所有参与联盟的平均边际贡献，公平地分配这些收益，即成员i最终分得的收益等于该成员为所加入的联盟创造的边际收益的平均值[8]。假设玩家的个数为M，则这些玩家能组成$2^{|M|}$个可能的子集。一个博弈可看成一个函数$v: 2^{|M|} \to \mathbf{R}$，每个子集都被映射到一个实数，第$i$个玩家的Shapley值如下

$$\psi(i|M)=\sum_{S\subseteq M\backslash\{i\}}\frac{|S|!(|M|-|S|-1)!}{|M|!}[a(S\cup\{i\})-a(S)] \tag{3}$$

式中：$|\cdot|$为集合的元素数量，$|M|$为整个集合，所有不包括玩家i的子集由$|S|$表示，$M\backslash\{i\}$代表从集合M中去掉玩家i后的子集。玩家i参与不同联盟S创造的边际贡献值记为$[a(S\cup\{i\})-a(S)]$。玩家的边际贡献值通过使用基线Shapley计算，即

$$a(S)=b(x_S;\tilde{x}_{N\backslash S}) \tag{4}$$

式中：$b(\cdot)$表示真实类别分类的得分，基线值$\tilde{x}$被设为0。在本研究中，将输入的SAR图像中的目标区域和背景区域视为两个玩家，并根据Shapley值计算其分别对模型识别率的贡献。

2 BDDCL模型算法

2.1 算法设计与分析

现有主流对比学习算法（如SimCLR[9]、MoCo[24]、BOYL[25]、SwAV[26]）在处理目标切片图像时，普遍将目标与背景作为一个整体进行表征学习。这种策略使得模型提取的特征不仅包含与下游分类任务相关的目标信息，还混杂了大量无关的背景信息，从而在SAR图像分类中易受背景干扰，影响识别精度和泛化能力。针对SAR图像中背景信息对下游分类任务的干扰，本文提出了一种基于双重对比学习的BDDCL模型算法。该算法在传统对比学习框架的基础上，新增了一个背景对比学习分支，旨在有效抑制背景信息对表征学习的影响。

BDDCL模型首先通过图像分割技术，将SAR图像分解为目标区域和背景区域。随后，通过将目标区域与随机选取的不同背景区域合成，构造出用于背景对比学习的样本：(1)锚点样本，即原始目标切片；(2)正样本集，通过背景变换生成的具有不同背景的目标样本；(3)负样本集，与锚点共享相同背景但目标不同的样本；(4)不同目标但背景相同的样本。这些样本共同参与特征学习过程，以实现目标信息与背景信息的分离。

背景对比学习分支采用对比学习的核心原则，即在特征空间中拉近锚点与正样本的距离，同时推远锚点与负样本的距离。具体而言，正样本因具备显著不同的背景，这一机制促使模型聚焦于目标区域的共有特征而非背景的多样性信息，从而降低了背景特征在表征中的占比。同时，尽管负样本与锚点共享相同背景，对比学习仍会增大两者之间的特征距离。此过程进一步削弱了背景信息的影响，使模型更关注目标的区分性特征。通过迭代训练优化，本算法逐步精炼特征表征，最终收敛时，学得的特征主要来源于目标区域。这一结果与传统对比学习模型的混合表征（目标与背景信息并存）形成鲜明对比。BDDCL通过新增的背景分支，实现了对背景信息的显式抑制和对表征学习的纠偏，为SAR图像分类任务提供了更纯净的目标特征。

2.2 算法整体流程

基于上述设计，本文提出的BDDCL模型表征提取流程如表1所示。算法通过双重对比学习机制，结合目标和背景分支，优化特征提取过程。

表1 BDDCL模型算法步骤

Table 1 BDDCL model algorithm step

BDDCL模型算法步骤
输入:批次大小N,样本集$X=\{x_1,x_2,\cdots,x_M\}$;样本集前景$F=\{f_1,f_2,\cdots,f_M\}$;样本集背景$B=\{b_1,b_2,\cdots,b_M\}$;温度系数τ,编码器$f_\theta(\cdot)$,投影头$g_\varphi(\cdot)$,数据增强组合T,负样本字典D
for一个训练批次的数据$\{x_k\}_{k=1}^{N}$ do
获得相应的前景图片$\{f_k\}_{k=1}^{N}$,背景图片$\{b_k\}_{k=1}^{N}$
获得两个数据增强函数t,t':$t\sim T,t'\sim T$
for all $k\in\{1,2,\cdots,N\}$ do
针对N张图片,分别进行两次数据增强,获得$2N$张视图:
第一次数据增强:$V_k=t(x_k)$, $\boldsymbol{h}_k=f(V_k)$, $\boldsymbol{z}_k=g(\boldsymbol{h}_k)$
V_k:第一次数据增强视图;$\boldsymbol{h}_k$:自注意力池化将输出激活展开得到的一维特征向量;$\boldsymbol{z}_k$:高维特征向量$\boldsymbol{h}_k$映射得到编码向量
第二次数据增强:$V_{k+N}=t(x_k)$, $\boldsymbol{h}_{k+N}=f(V_{k+N})$, $\boldsymbol{z}_{k+N}=g(\boldsymbol{h}_{k+N})$
V_{k+N}:第二次数据增强视图;$\boldsymbol{h}_{k+N}$:自注意力池化将输出激活展开得到的一维特征向量;$\boldsymbol{z}_{k+N}$:高维特征向量$\boldsymbol{h}_{k+N}$映射得到编码向量
根据背景标签将z_k^{pos}加入负样本字典D:
$V_k^f=t(f_k)$, $\boldsymbol{V}_k^{\text{pos}}=V_k^f+b_{\text{index}}(\text{index}\neq k)$, $\boldsymbol{h}_k^{\text{pos}}=f(\boldsymbol{V}_k^{\text{pos}})$, $\boldsymbol{z}_k^{\text{pos}}=g(\boldsymbol{V}_k^{\text{pos}})$
V_k^f:前景图片的第一次数据增强视图; V_k^{pos}:前景增强视图与不同背景图片构造样本视图; $\boldsymbol{h}_k^{\text{pos}}$:自注意力池化将输出激活展开得到的一维特征向量;$\boldsymbol{z}_k^{\text{pos}}$:高维特征向量$\boldsymbol{V}_k^{\text{pos}}$映射得到编码向量
根据背景标签将z_{k+N}^{pos}加入负样本字典D':
$V_{k+N}^f=t'(f_{k+N})$, $\boldsymbol{V}_{k+N}^{\text{pos}}=V_{k+N}^f+b_{\text{index}}(\text{index}\neq k+N)$, $\boldsymbol{h}_{k+N}^{\text{pos}}=f(\boldsymbol{V}_{k+N}^{\text{pos}})$, $\boldsymbol{z}_{k+N}^{\text{pos}}=g(\boldsymbol{V}_{k+N}^{\text{pos}})$
V_{k+N}^f:前景图片的第二次数据增强视图; V_{k+N}^{pos}:前景增强视图与不同背景图片构造样本视图; $\boldsymbol{h}_{k+N}^{\text{pos}}$:自注意力池化将输出激活展开得到的一维特征向量;$\boldsymbol{z}_{k+N}^{\text{pos}}$:高维特征向量$\boldsymbol{V}_{k+N}^{\text{pos}}$映射得到编码向量
end for
计算InfoNCE实例级判别损失,即$\mathcal{L}_{\text{con}}^1$分支对比损失:
$$\mathcal{L}_{\text{con}}^1=-\sum_{i=1}^{2N}\log\frac{\exp(z_i\cdot\boldsymbol{z}_p^{\mathrm{T}}/\tau)}{\sum_{k=1,k\neq i}^{2N}\exp(z_i\cdot\boldsymbol{z}_k^{\mathrm{T}}/\tau)}$$
从负样本字典D中取出与$z_i(i\in[1,N])$具有相同背景标签的负样本d_i,数量为A,计算$\mathcal{L}_{\text{con}}^2$分支对比损失:
$$\mathcal{L}_{\text{con}}^2=-\sum_{i=1}^{N}\log\frac{\exp\left(z_i\cdot(\boldsymbol{z}_i^{\text{pos}})^{\mathrm{T}}/\tau\right)}{\sum_{k=1}^{A}\exp(z_i\cdot\boldsymbol{d}_k^{\mathrm{T}}/\tau)}$$
从负样本字典D'中取出与$z_i(i\in[N+1,2N])$具有相同背景标签的负样本d_i,数量为A,计算$\mathcal{L}_{\text{con}}^3$分支对比损失:
$$\mathcal{L}_{\text{con}}^3=-\sum_{i=N+1}^{2N}\log\frac{\exp\left(z_i\cdot(\boldsymbol{z}_i^{\text{pos}})^{\mathrm{T}}/\tau\right)}{\sum_{k=1}^{A}\exp(z_i\cdot\boldsymbol{d}_k^{\mathrm{T}}/\tau)}$$
计算最终损失函数:$\mathcal{L}=(\mathcal{L}_{\text{con}}^1+\mathcal{L}_{\text{con}}^2+\mathcal{L}_{\text{con}}^3)/3$
更新编码器$f_\theta(\cdot)$与投影头$g_\varphi(\cdot)$来最小化损失$\mathcal{L}$
end for
输出:编码器$f_\theta(\cdot)$用于提取数据集特征向量进行下游任务,并丢弃投影头$g_\varphi(\cdot)$

3 实验结果与分析

3.1 数据集与实验环境

本文基于MSTAR数据集评估BDDCL模型性能，所有实验均在标准操作条件(Standard operating conditions，SOC)[4]下进行。为确保结果一致性，严格控制了俯仰角、信噪比和分辨率等变量：训练集采用17°俯仰角数据，测试集采用15°俯仰角数据，信噪比均大于30 dB，分辨率为0.3 m。为保证对比实验的公平性，训练与测试集在同一目标类别中选用相同型号，其中BMP-2和T-72分别采用BMP2-9563和T72-132型号构建训练集，以减少无关变量的干扰。表2列出目标类型、序列号、俯仰角及样本数等数据集的详细信息。

本文的验证实验是在一台高性能GPU工作站上完成，其硬件配置为：Intel(R) Xeon(R) Sliver 4215R CPU (3.20 GHz)、256.0 G RAM和NVIDIA GeForce RTX 3090 GPU。工作站上安装有64位Windows10系统，并安装了python 3.7和深度学习工具包PyTorch 1.10.2。所有模型训练与测试均在此环境下基于Python和PyTorch实现。

表2 MSTAR数据集详细信息

Table 2 Detailed information about the MSTAR dataset

版本	序列号	样本数量	
		17°俯仰角	15°俯仰角
2S1	b01 9563	299	274
		233	194
BMP2	9566 c21	232	196
		233	196
BRDM2	E-71	298	274
BTR60	K10yt7532	256	195
BTR70	c71	233	196
D7	92v13015	299	274
T62	A51 132	299	273
		232	196
T72	812 s7	231	195
		233	191
ZIL131	E12	299	274
ZSU234	d08	299	274

3.2 模型性能评估与分析

通过分类性能、特征贡献分析、可视化验证及运行效率等多维度评估，系统分析了BDDCL模型在SAR图像分类任务中的表现。实验结果表明，BDDCL模型不仅在SOC条件下实现了高精度目标识别，还有效抑制了背景伪相关性，并在运行效率上表现优异，充分验证了其在复杂场景下的适用性。

为全面评估BDDCL模型的分类性能，利用训练好的编码器从图像中提取特征后，采用SoftMax分类器[27]进行性能测试，并使用定义为正确识别的样本数与总样本数的比值的正确分类概率(P_{cc})作为评价模型性能的指标。表3展示了BDDCL模型在SOC下的识别结果混淆矩阵，矩阵对角线表示各目标

表3 BDDCL模型在标准操作条件下的识别结果

Table 3 Recognition results of BDDCL model under SOC

类别	2S1	BMP2	BRDM2	BTR60	BTR70	D7	T72	T62	ZIL131	ZSU234	P_{cc}/%
2S1	258	0	12	0	0	1	1	0	2	0	94.16
BMP2	1	189	0	0	1	0	2	2	0	0	96.92
BRDM2	1	2	267	0	1	0	0	0	0	3	97.44
BTR60	1	0	5	181	0	0	0	0	0	0	96.79
BTR70	6	1	8	0	181	0	0	0	0	0	92.35
D7	0	0	0	0	0	271	0	0	3	0	98.91
T72	8	0	1	1	0	0	252	0	3	1	94.74
T62	0	0	0	1	0	0	0	194	0	1	98.98
ZIL131	0	0	3	0	0	4	0	0	230	0	97.05
ZSU234	2	0	2	2	0	0	6	0	1	261	95.26
Overall											96.24

类别正确识别的样本数。结果显示总体识别率达96.24%,具体而言,T62和D7的识别率超过98%,BRDM2和ZIL131超过97%,其余类别均在92%以上。这表明BDDCL模型能够从10类目标中提取具有区分性的稳健特征,在原始数据集上展现出卓越的分类性能。模型训练损失曲线如图5所示,进一步反映了训练过程的稳定性。

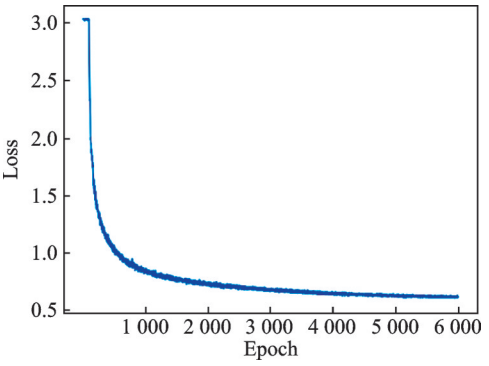


图5 BDDCL模型训练损失曲线

Fig.5 BDDCL model training loss curve

为探究上述高性能背后的特征学习机制并分析BDDCL模型对伪相关背景信息的抑制能力,计算了SAR图像目标和背景区域在训练过程中Shapley值的变化,结果如图6所示。由图可知,目标区域的Shapley值在前1 500轮显著上升,至3 500轮后稳定在0.6左右;而背景区域的Shapley值始终维持在0.1左右,远低于目标区域。这表明目标区域对模型特征学习的贡献显著高于背景,模型主要依赖目标信息进行分类决策,而背景信息的影响微乎其微,与目标类别间的相关性极低,从而有效抑制了伪相关性。

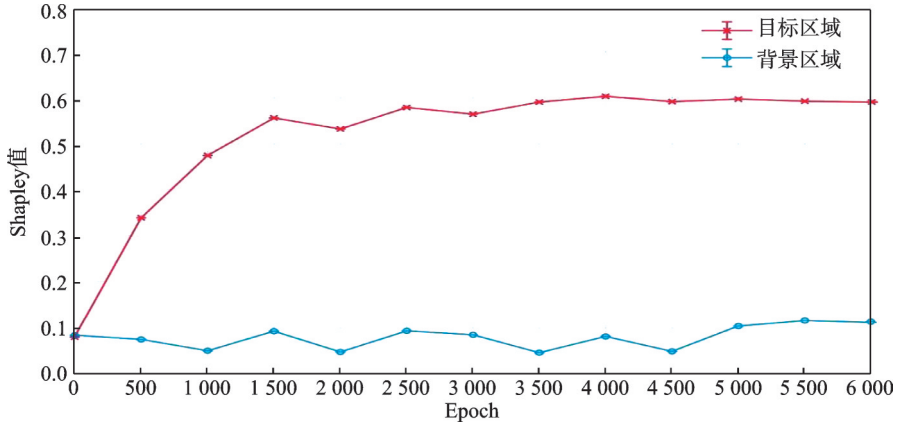


图6 BDDCL模型训练过程中SAR图像目标区域及背景区域的Shapley值曲线

Fig.6 Shapley value curves of target area and background area of SAR images during BDDCL model training process

鉴于Shapley值分析提示了背景抑制效果,为直观验证所提方法在缓解背景表征偏差问题上的作用,采用网络可视化技术,输出经ResNet50第4个Basic block处理的特征图,如图7所示。特征图通过亮度反映网络对图像区域的激活强度,亮度越高表示激活值越大。图7(a)显示未纠偏时特征图整体偏

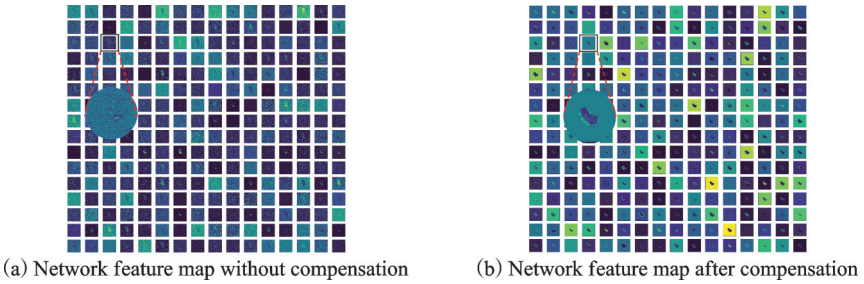


图7 背景纠偏前后网络特征图

Fig.7 Feature maps of the network before and after background debiasing

暗,网络在背景区域的激活值较高,对目标关注不足,导致背景信息被过度学习;图7(b)显示经背景纠偏后,特征图在目标区域的激活值显著增强,而背景区域几乎无激活。这表明所提方法有效降低了背景表征偏差,提升了模型对目标区域的关注能力。

为进一步评估BDDCL模型的运行效率,统计了训练和预测阶段的耗时。训练阶段使用128像素×128像素的样本,总数为2 687,batch size设为16,平均完成一个训练周期耗时约1 min。测试阶段样本大小同为128像素×128像素,总数为2 373,batch size为16,预测总耗时约31 s,平均单样本预测时间为13.06 ms。这些结果表明,模型在保持高性能的同时,具有较高的运行效率。

3.3 模型对比与泛化验证

为了全面评估所提出的BDDCL模型的性能,通过与传统的对比学习模型SimCLR[9]和MoCo[24]的对比,分析了在识别性能、目标区域贡献度及背景伪相关抑制能力上的表现差异。实验在SOC下进行,结果如表4所示。BDDCL模型的识别率为96.24%,分别较SimCLR模型(96.16%)和MoCo(96.08%)提升0.08%和0.16%,表明其在SAR图像分类中具有优越性。进一步基于Shapley值评估目标区域贡献度,BDDCL占比达84.0%,比SimCLR模型(64.7%)和MoCo模型(63.5%)分别提高19.3%和20.5%,而背景区域占比显著降低。这验证了BDDCL模型能有效增强目标特征表达并抑制非因果背景干扰,展现出对多种正负样本对比学习框架的普适性和背景纠偏能力。

表4 标准操作条件下实验结果

Table 4 Experimental results under SOC

参数	BDDCL模型	SimCLR模型	MoCo模型
原始图像的P_{cc}/%	96.24	96.16	96.08
目标区域的Shapley值	0.593	0.321	0.303
目标区域的Shapley值百分比/%	84.0	64.7	63.5
背景区域的Shapley值	0.113	0.175	0.174
背景区域的Shapley值百分比/%	16.0	35.3	36.1

为验证BDDCL模型对陌生背景的泛化能力,设计了新测试集,保留原始图像的目标区域并将背景替换为随机类别背景,用以测试BDDCL模型和SimCLR模型的分类性能。如表5所示,实验结果显示,BDDCL模型在陌生背景下的总体识别率达94.35%,其中T62识别率达100%,BTR60、D7、ZIL131和ZSU234均超过95%,表明其能稳定提取目标特征,对背景变化表现出显著的适应性。相比之下,

表5 BDDCL模型对于陌生背景下图片的识别结果

Table 5 Recognition results of BDDCL model for pictures under unfamiliar background

类别	2S1	BMP2	BRDM2	BTR60	BTR70	D7	T72	T62	ZIL131	ZSU234	P_{cc}/%
2S1	258	0	3	0	0	1	7	0	5	0	94.16
BMP2	0	187	0	0	0	0	0	8	0	0	95.89
BRDM2	24	0	227	2	4	3	6	0	8	0	82.84
BTR60	0	1	4	178	0	0	1	0	3	0	95.18
BTR70	3	12	0	3	176	0	0	2	0	0	89.79
D7	0	0	0	0	0	273	0	0	1	0	99.63
T72	3	0	1	1	0	0	248	1	11	1	93.23
T62	0	0	0	0	0	0	0	196	0	0	100
ZIL131	0	0	0	0	0	4	0	0	233	0	98.31
ZSU234	0	0	1	0	0	3	3	0	4	263	95.98
Overall											94.35

SimCLR模型识别率仅为54.36%(见表6),性能明显逊于BDDCL模型。分析表明,传统对比学习方法因训练时过度依赖背景特征,在面对陌生背景时鲁棒性不足;而BDDCL模型通过聚焦目标区域并有效抑制背景伪相关性,在复杂背景中依然保持高精度分类,凸显了其优异的泛化能力和鲁棒性。

表6 陌生背景下图片的识别结果

Table 6 Recognition results of images under unfamiliar backgrounds

模型	P_{cc}/%
SimCLR	54.36
BDDCL	94.35

4 结束语

本文针对SAR ATR深度表征学习中背景表征的偏差问题,提出了一种基于双重对比学习模型的背景伪相关去偏方法。通过背景变化的数据增强技术构造了解耦目标与背景的正负样本对。该方法不仅比较不同SAR图像目标区域,也对背景区域进行了对比,从而鼓励模型聚焦于目标前景的语义内容,并惩罚从无关背景区域提取特征。实验结果和通过Shapley值进行的可解释性分析证明了所提方法能够在保持表征性能的同时,有效减少模型对无关背景信息的过拟合,提出的背景去偏模型显著提升了对变化背景的鲁棒性,更符合实际应用需求。

参考文献:

[1] 罗汝, 赵凌君, 何奇山, 等. SAR图像飞机目标智能检测识别技术研究进展与展望[J]. 雷达学报, 2024, 13(2): 307-330.
LUO Ru, ZHAO Lingjun, HE Qishan, et al. Intelligent technology for aircraft detection and recognition through SAR imagery: Advancements and prospects[J]. Journal of Radars, 2024, 13(2): 307-330.

[2] QI G J, LUO J. Small data challenges in big data era: A survey of recent progress on unsupervised and semi-supervised methods [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(4): 2168-2187.

[3] WANG C, GU H, SU W. SAR image classification using contrastive learning and pseudo-labels with limited data[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.

[4] CHEN S, WANG H, XU F, et al. Target classification using the deep convolutional networks for SAR images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(8): 4806-4817.

[5] 向卫力, 李晓辉, 周勇胜, 等. 一种鲁棒的多尺度稀疏表示SAR目标识别方法[J]. 中国科学院大学学报, 2017, 34(1): 99-105.
XIANG Weili, LI Xiaohui, ZHOU Yongsheng, et al. A robust SAR target recognition method based on multi-scale feature and sparse representation[J]. Journal of University of Chinese Academy of Science, 2017, 34(1): 99-105.

[6] WANG L, BAI X, ZHOU F. SAR ATR of ground vehicles based on ESENet[J]. Remote Sensing, 2019, 11(11): 1316.

[7] YANG L, LU L, LIU C, et al. Interactive exploration of CNN interpretability via coalitional game theory[J]. Scientific Reports, 2025, 15(1): 9261.

[8] ZHANG H, XIE Y, ZHENG L, et al. Interpreting multivariate Shapley interactions in DNNs[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI, 2021: 10877-10886.

[9] BELLONI C, AOUF N, BALLERI A, et al. Explainability of deep SAR ATR through feature analysis[J]. IEEE Transactions on Aerospace and Electronic Systems, 2021, 57(1): 1-10.

[10] 李玮杰, 杨威, 刘永祥, 等. 雷达图像深度学习模型的可解释性研究与探索[J]. 中国科学:信息科学, 2022, 52(6): 1114-1134.
LI, Weijie, YANG Wei, LIU Yongxiang, et al. Research and exploration on the interpretability of deep learning models for radar images[J]. Scientia Sinica Informationis, 2022,52(6):1114-1134.

[11] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]//Proceedings of International Conference on Machine Learning. Los Angeles: PMLR, 2020: 1597-1607.

[12] HEILIGERS M, HUIZING A. On the importance of visual explanation and segmentation for SAR ATR using deep learning [C]//Proceedings of 2018 IEEE Radar Conference (RadarConf18). Oklahoma City: IEEE, 2018: 0394-0399.

[13] LI C, DU L, DENG S, et al. Point-wise discriminative auto-encoder with application on robust radar automatic target

recognition[J]. Signal Processing, 2020, 169: 107385.
[14] WANG L, BAI X, ZHOU F. SAR ATR of ground vehicles based on ESENet[J]. Remote Sensing, 2019, 11(11): 1316.
[15] FENG S, JI K, WANG F, et al. PAN: Part attention network integrating electromagnetic characteristics for interpretable SAR vehicle target recognition[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-17.
[16] LEWIS B, DEGUCHY O, SEBASTIAN J, et al. Realistic SAR data augmentation using machine learning techniques[C]// Proceedings of Algorithms for Synthetic Aperture Radar Imagery XXVI. San Francisco: SPIE, 2019, 10987: 1-12.
[17] 侯兴荣, 彭冲. 基于局部相似性学习的鲁棒非负矩阵分解[J]. 数据采集与处理, 2023, 38(5): 1125-1141.
HOU Xingrong, PENG Chong. Robust non-negative matrix factorization based on local similarity learning[J]. Journal of Data Acquisition and Processing, 2023, 38(5): 1125-1141.
[18] PENG B, XIE J, PENG B, et al. Learning invariant representation via contrastive feature alignment for clutter robust SAR ATR[J]. IEEE Geoscience and Remote Sensing Letters, 2023, 20(20): 1-5.
[19] LI W, YANG W, ZHANG W, et al. Hierarchical disentanglement-alignment network for robust SAR vehicle recognition[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023, 16(16): 9661-9679.
[20] VAN DEN OORD A, LI Y, VINYALS O J A E P. Representation learning with contrastive predictive coding[J]. arXiv preprint arXiv: 1807.03748, 2018.
[21] 赵明波, 何峻, 付强. SAR图像CFAR检测的快速算法综述[J]. 自动化学报, 2012, 38(12): 1885-1895.
ZHAO Mingbo, HE Jun, FU Qiang. Survey on fast CFAR detection algorithms for SAR image targets[J]. Acta Automatica Sinica, 2012, 38(12): 1885-1895.
[22] SHAFIQ M, GU Z. Deep residual learning for image recognition: A survey[J]. Applied Sciences, 2022, 12(18): 8972.
[23] JING L, TIAN Y. Self-supervised visual feature learning with deep neural networks: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(11): 4037-4058.
[24] HE K, FAN H, WU Y, et al. Momentum contrast for unsupervised visual representation learning[J]. arXiv preprint arXiv: 1911.05722, 2020.
[25] GRILL J B, STRUB F, ALTCHÉ F, et al. Bootstrap your own latent—A new approach to self-supervised learning[J]. Advances in Neural Information Processing System, 2020, 33: 21271-21284.
[26] CARON M, MISRA I, MAIRAL J, et al. Unsupervised learning of visual features by contrasting cluster assignments[J]. Advances in Neural Information Processing System, 2020, 33: 9912-9924.
[27] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep learning[M]. Cambridge: MIT Press, 2016.

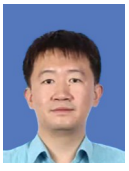
作者简介:

张文青(1969-),男,实验师,研究方向:电子测量、智能仪器、信号分析与处理,E-mail: zhangwq@njust.edu.cn。

王景(2000-),女,硕士研究生,研究方向:合成孔径雷达图像处理。

黄雪琴(2001-),**通信作者**,女,硕士研究生,研究方向:合成孔径雷达遥感与目标识别,E-mail:123104024862@njust.edu.cn。

田巳睿(1981-),男,副教授,研究方向:合成孔径雷达遥感与目标识别、声信号处理与识别。



何成(1980-),男,高级工程师,研究方向:雷达系统总体及科技信息领域研究。

张劲东(1981-),男,副教授,研究方向:认知雷达与智能抗干扰、雷达信号处理等。

李洪涛(1979-),男,副教授,研究方向:雷达信号处理、阵列信号处理设计及应用、高速数字信号系统设计等。

(编辑:夏道家)